Auditor Screening in the Presence of Social Polarization

Abstract

Polarization has been intensifying in recent years and manifested in various facets of our daily lives. What was initially a phenomenon on the ideological political spectrum ("ideological polarization") has since become a strongly emotional one based on one's social ties ("affective polarization"), influencing individual and group actions largely based on their group identity ("social polarization"). Intuitively, due to its adversarial nature, one would surmise that accounting audits are particularly vulnerable to social polarization, with auditors' decisions influenced, whether consciously or unconsciously, by their clients' group identity on the polarization spectrum. We first investigate population-level network formation mechanisms that closely model the polarization seen empirically, particularly in the United States. In doing so, we pay a particular close attention to how polarization undergoes a phase change from ideological to affective, which has significant implications for adversarial transactions such as audits. We then discuss the problem of auditor screening from the perspective of an impartial audit committee in a dynamic setting, where the pool of auditors is a sample from the population with a given level of polarization. Auditors ("agents"), just like other members of the society ("individuals"), dynamically form their social networks to minimize cognitive dissonance that arises from the mismatch between actions implied by their own ideologies and circumstances, and those of their friends, family, and other members of their social network. We find that in equilibrium, given that individuals place sufficient weight on the behaviors of their peers when optimizing their actions (the affective polarization parameter), disparate network communities emerge that partition the network and action space. As a result, the group identity, rather than experiences or ideologies, of the auditor determines their action particularly in situations without clear-cut answers. We then characterize the class of auditor selection mechanisms that are optimal for impartial audit committees. Given the CFO's position on the polarization spectrum, the optimal auditor is one whose social network is the most homogeneously polarized on the opposite end of the spectrum.

1 Introduction

Political polarization has increased in recent years, with people in different corners of the political spectrum growing increasingly divided ideologically and, more importantly, emotionally. It has spawned active multidisciplinary research. In political science literature, it is widely held that those who are most ideologically oriented are also the most politically active, often referred to as "political elites"(Enders, 2021; Callander and Carbajal, 2022). They tend to adopt extreme ideological positions to cater to the mass public to pull them closer to their end of the spectrum (Enders, 2021). Layman et al. (2006) noted that political activists are a main spring of elite polarization due to their considerable influence in shaping the ideologies of the masses. However, there is increasing evidence that suggests that in contrast to the elites, polarization among the mass public has been characterized by emotion (or affective)-driven with their group identity largely driving their positions on ideological issues and actions (Baldassarri and Page, 2021).

This trend has wide-reaching implications for corporate financial audits. As members of the polarized mass public, participants in the audit process are subject to polarizing influences. (Felix et al., 2024; Bhandari et al., 2020). Such biases can be ideological and, more importantly, emotional in nature and there is increasing evidence pointing to the influence that these (unconscious) biases on audit processes and participants. Felix et al. (2024) observed that political dissimilarity between the CFO and audit committee can improve audit quality. Bhandari et al. (2020) presented evidence that supports the claim that Republican CEOs are associated with lower inherent and control risk. Hasan and Jiang (2022) observed that strength in corporate social responsibility (CSR) was positively associated with political sentiment in firms' filings, (Arikan et al., 2022) noted that CEOs whose partisan identities aligned with US presidents' expressed greater optimism in financial forecasts and disclosures, and (Fos et al., 2022) showed a significant rise in assortative matching along political party lines among the executive ranks. Partisan alignment was linked to executives' subjective perception of future economic uncertainty in everyday decision making (Ambrocio and Hasan, 2022). Politics introduces biases in the industries that depend on a favorable political climate, such as access to natural resources and government contracts (Boubakri et al., 2012). Executives' biases from firm ideologies can lead to less compromise during meetings with politically-misaligned executives more likely to leave and the remaining team tending to recruit party-aligned executives.

Literature offers two main explanations for how the mass public becomes polarized. Ideological polarization occurs when individuals congregate based on their rational assessment of the similarity between their stances on relevant policy issues (Hetherington, 2001; Baldassarri and Page, 2021). For instance, Callander and Carbajal (2022) argue that polarization occurs on the ideological spectrum through elections where political parties representing the "elites" and the mass public coalesce into polarized groups. Polarization can also be grounded in emotions, also called "affective polarization" (Baldassarri and Page, 2021; Druckman et al., 2013; Cohen, 2003; Theodoridis, 2017), in which an agent develops a strong affinity for in-group members while harboring animosity towards outsiders. As a result, individuals place greater trust in opinions of their own group, even those less substantively grounded (Druckman et al., 2013). In an influential book, Mason (2018) refers to this emotion-driven polarization among the masses due to the influence of their social ties "social polarization".

Evidence suggests ideological polarization occurs first among elites, followed by affective polarization among non-elites,(Callander and Carbajal, 2022), resulting in a bipolar segregation of the mass public with strong in-group emotional attachment and out-group animosity.(Baldassarri and Page, 2021; McCoy et al., 2018; Dias and Lelkes, 2022). Callander and Carbajal (2022) argue that "Polarization is a dynamic path with elites polarizing first and more dramatically, mass polarization coming later less pronounced". Baldassarri and Page (2021) state that "the type of polarization in the US is affective not ideological; the former is 'fueled by emotional attachment and repulsion"'. Dias and Lelkes (2022) noted that "The partisan divide on Covid policy has been widely reported, and voters more driven by emotions and less by political ideologies." In addition, a stream of studies has investigated the association between affective polarization and individual behavior (McCoy et al., 2018; Baldassarri and Page, 2021).

In this paper, we study the problem of screening auditors in the presence of social polarization, which is readily applicable to general adversarial transactional settings. To situate auditor selection in the framework of varying degrees of social polarization, we first investigate a population-level discrete-time network formation model that delineate behavioral mechanisms that lead to different forms of polarization in a society. Then, we examine its impact on auditor selection as an example of a more general screening problem against adverse selection. Our research sheds light on the optimal auditor selection mechanism that arises as a result and policy implications for regulatory bodies like the Public Company Accounting Oversight Board.

First, we propose a dynamic network games model to explain empirical phenomena documented in recent literature on polarization among executives and the society at large, and depict various concepts of equilibrium and stability based on the best-response framework. While political elites are ideologically polarized (Callander and Carbajal, 2022; Clarke et al., 2021; Dias and Lelkes, 2022), we show that board members ("agents"), along with the rest of the mass public, become polarized increasingly through the "affective social" channel comprising peer effects into disparate communities, each characterized by uniform ideologies and, thus, actions. Ideologies and affective motivations of agents are manifested through their actions on a set of issues that the group is tasked to make decisions on. The actions of each agent are influenced by a combination of three key factors: (1) time-varying their own ideologies, encompassing their personal beliefs on a variety of issues; (2) time-invariant personal circumstances, such as educational background and career trajectory; and (3) their social network, which exerts an increasingly influential control over their actions through the affective framing of "in-group vs. out-group" (Tversky and Kahneman, 1981), particularly as the network becomes more polarized. This interplay between individual ideologies, personal circumstances, and the dynamics of the social network underscores the intricate mechanisms that drive and amplify polarization effects. Agents wish to form relationships with those possessing similar ideologies and circumstances. However, since one's ideologies and certain circumstances are not visible to others, agents infer each other's intangibles by observing their actions. Relationships are consummated when both agents involved simultaneously consent. The resulting network is formed endogenously in discrete time steps, and is undirected and exhibits homophily.

To build intuition, we investigate the evolution of network structures with fixed initial symmetric ideology distribution. Exogenous shocks to ideologies of elites polarize their neighbors both ideologically and emotionally over time. We consider a group of business executives, denoted as "business elites", who work closely with polarized political elites. In real-world settings, these executives could include lobbyists and CEOs in industries with significant government presence, such as defense and telecommunications. We show that business elites in turn affect the ideologies and actions of those in their networks, and increasingly visible signs of political polarization ensue in the executive circles and the general public. We show that in the absence of exogenous shocks to ideologies, the distribution of actions is uniform with all agents belonging to one giant connected component whose ideologies converge to the moderate median. However, with exogenous shocks affecting even a small minority of the executive members, disconnected network communities emerge, exacerbating polarization without communication between them.

2 Related Literature

This paper builds on several strands of research. Formally, the subject of our investigation is dynamic coordination game on networks with strategic complementarities under the assumption of bounded rationality. Our framework builds on graphical games originally studied in computer science where agent interactions on networks are analyzed as games (Kakade et al., 2003; Kearns et al., 2013), and “semi-anonymous graphical games” (Galeotti et al., 2010; Orlova, 2022)[1]. While in these models, an individual’s behavior is determined statically based on their equal-weighted friends, in our model, an agent behavior depends not only on their neighbors, but also on their latent ideologies and characteristics, and their interactions are dynamic. In this sense, our approach differs from coordination games. In addition, while identities of neighbors have no role in semi-anonymous games, we make explicit their role with heterogeneous weights ρ_i.

“Stubborn agents" in Acemoğlu et al. (2013) bear some resemblance to our “elite executives” in that they do not sway from their opinions. However, elites in our model have an explicit ultimate goal of maximizing the size of their neighborhood by strategically approaching agents with ideologies most similar to theirs. There is a stream of research focused on strategic network formation.

Bramoullé et al. (2014) discovered that algebraic spectral properties of networks can be informative in predicting Nash and stable equilibria. Although the authors incorporated strategic aspects, the network formation considered there proceeded exogenously based on one key model parameter: the payoff constant that determines the marginal cost of link formation. Bolletta and

[1] See Appendix for a definition and references for a detailed treatment on this class of games.

Pin (2020) investigated endogenous network formation and initial conditions that result in polarization by assuming that an agent's action is a convex combination of hers and the mean of her neighbors, with the objective of minimizing this difference. Their agents are semi-anonymous and exhibit "naive learning" (Golub and Sadler, 2017) in a sense that they use only the information about their one-dimensional opinions. In contrast, we model agents as unique members of two classes and exhibit a more robust form of bounded rationality by incorporating potentially a large number of factors and history when deciding on forming links with other players. In characterizing equilibrium concepts, we build on the notions of equilibrium viability Kalai (2020); Kalai and Kalai (2021); Kim et al. (2022) by extending to network games and refining the notions of defection-deterrence and formation difficulty (Kalai, 2020) for equilibrium networks.

More recently, Parise and Ozdaglar (2023) put forward a novel approach by introducing an infinite-player generalization of network games, which serves as a limiting case, to address the computational complexity and the analytical tractability of large network games by capitalizing on the inherent low-rank structure of the network generative model known as graphons. However, their primary focus was on establishing the asymptotic convergence of large network games towards graphon games, rather than delving into the dynamics of network polarization or its consequential impact on the agent decision-making.

3 The Network Game

We first model the mechanism for polarization of the population. We use the term "society" to denote the population. We walk through how our model can explain various phenomena relating to polarization as observed empirically in the literature and the popular press, and set context for delineating the impact of various forms of polarization on auditor selection and policy implications. While network games studied in the literature typically assume a setting with a finite number of players in finite parametric dimensions, we generalize such a setup to the countably-infinite player setting in infinite parametric dimensions using tools from Hilbert spaces. Doing so greatly simplifies proofs for analytical results and interpretations. We find all Nash and stable equilibria under our network games structure and parametric assumptions. We start with the basic structure of the network game, followed by a discussion of how different notions of polarization is incorporated into our model[2].

3.1 Preliminaries

$\langle x, y\rangle$ denotes the inner product in the Hilbert space $\mathcal{H} := \ell^2(\mathbb{R}), x, y, \in \mathcal{H}$, and $\odot$ the component-wise multiplication. The norm induced by the inner product is denoted as $\langle x, x\rangle := \|x\|$. We use italics to refer to vectors and boldface capital letters to refer to operators. $\mathbf{1}^k$ denotes the k-vector of all 1s, $\mathbf{1}^k_a$ denotes the k-vector of a's in all entries, and $\mathbf{I}_A(x)$ denotes the indicator that evaluates to 1 if $x \in A$ and 0 otherwise. For any vector v, v_{-i} denotes v with the i-th component removed and v_S refers to a subset of v defined by $\{j : j \in S \cap v\}$. For any $N \in \mathbb{N}$, $[N]$ represents the set of natural numbers $\{1, 2, ..., N\}$. $\#S$ denotes the cardinality of a finite countable set S and $|T|$ represents the Lebesgue measure of an uncountable set T.

Undirected graph is denoted $\mathbf{G} := (\mathbf{V}, \mathbf{E})$ with the node set $\mathbf{V}$ with $|\mathbf{V}| \leq \infty$ representing agents and the edge set $\mathbf{E}$ representing relationships between pairs of agents. Subsets of $\mathbf{V}$ are denoted coalitions of agents and $\mathbf{V}$ is referred to as the grand coalition. For any set S, we denote the set difference by $S \setminus s := \{t : t \in S, t \neq s\}$. We represent all graphs as adjacency matrices $\mathbf{G}$ with $\mathbf{G}_{ij} = \mathbb{I}_{ij\in\mathbf{E}}$ where $\mathbb{I}_x$ is the Kronecker delta that equals 1 (respectively, 0) if the condition x evaluates to true (respectively, false). By definition of adjacency matrices, $\mathbf{G}_{ii} = 0$ for all i, i.e., we only consider relationships between two distinct agents. Using this notation, the set of friends for agent i is simply the i-th row of $\mathbf{G}$ and is denoted $\mathbf{G}_{i\cdot}$, and their sum (finite by definition of $\mathcal{H}$) is the degree of row i and denoted $d_i := |\mathbf{G}_{i\cdot}| = |\mathbf{G}_{\cdot i}|$. The terms "graph" and "network" are used interchangeably in this paper and both refer to $\mathbf{G}$.

[2]The longer version of this paper includes an appendix which contains definitions for several preliminary concepts and results from the network games literature that our results use, as well as a primer on operator theory and Hilbert spaces.

3.2 The game setup

Let subscripts $i, j \in [n]$ denote typical "individuals" of a society $\mathbf{V}$, and suppose that each agent serves on one or more boards of directors. At time 0, nature exogenously endows each i with time-invariant characteristics $X_i^0 \in \mathcal{H}$, such as place of birth, education, and career history, and ideologies $Y_i^0 \in \mathcal{H}$, such as partisan identity. However, due to individual's cognitive constraints (Cowan et al., 2005), at any given time t, individuals consider only a certain finite number K elements of X_i^t and Y_i^t when deciding on an action. For simplicity, suppose X_i^t and Y_i^t are sampled from a symmetric distribution[3] and known only privately to i. Each individual is identified with a node in a network where links between individuals reflect relationships.

Let convex weights $\psi_i^t, \phi_i^t \in \mathcal{H}_+ := \ell^2(\mathbb{R}_+)$ denote the relative importance that i places on the elements in X_i^t and Y_i^t, respectively, and $\rho_i^t \in \mathcal{H}_+$ the weights that i places on actions of other individuals they consider when deciding on actions. However, cognitive constraints imply $\#\{\psi_i^t, \phi_i^t > 0\} = K < \infty$ and similarly $\#\{\rho_i^t > 0\} < \infty$. We call the set of individuals for whom $\rho_i^t > 0$ individual i's *community of neighbors* and denote it by N_i^t with $\#N_i^t = n_i^t$.

At $t = 1$ individuals take two actions in sequence. First, each individual ("source") decides which other individuals ("target") she wants to be linked to. A link is established if both individuals voluntarily consent. Then, she decides which action to take on some k tasks. While the action space is infinite, due to the principle of cognitive constraint from behavioral economics (Tyson, 2008; Whitney et al., 2008), individuals take action on a finite number $k < \infty$ of tasks. Notice then that $a_i^t \in \mathcal{H}$ for all t as are X_i^t and Y_i^t. We formally define players' actions below.

Definition 3.1. *In each period t in the game, individual $i, \forall i$, takes two actions sequentially:*

Stage 1: *Names individuals to request to form links with, each at a cost of $c_f > 0$.*

(1) *Whether to maintain the link with individual $j, \forall j \in N_i^t$*

(2) *Whether to initiate a link with an individual not in her community, i.e., $l \in N_i^t \ \forall l \notin N_i^t$*

[3] We assume a symmetric distribution for tractability here. It can be shown that our results generalize under other distributions as well.

Stage 2: *Takes action on each of K tasks, each at a social pressure cost $c_a : \mathcal{H} \to \mathbb{R}_+$;*

All individuals play Stage 1 and influence each other's payoff in a manner akin to local public goods game in Bergstrom et al. (1986) in that neighbors affect each other's payoff. More specifically, the more of an agent's neighbors take actions identical to the agent, the higher the network's contribution to her payoff, making our game setup one of strategic complementarities (Galeotti et al., 2010). The social pressure cost c_a is concave and increasing in the difference between an agent's action and those of peers as measured by their inner product in the case of the binary action space and the induced norm in the case of the continuum action space. The social networking cost $c_f : \mathbb{N} \to \mathbb{R}^+$ is concave and increasing in the degree of row i, and represents both the cost incurred in maintaining an existing as well as initiating a new link. $c_f(x)$ accounts for time spent engaging in social activities to form or maintain relationships with x neighbors. We take c_f and c_a to be normalized and invariant across time, i.e., $0 = c_f(0) < c_f(n) < n$ and $0 = c_a(0) < c_a(k) < k$ for all $t, n \in \mathbb{N}$. As we will see later, $c_f > 0$ implies that individuals form links with those that are sufficiently similar to them in terms of ideologies and characteristics.

First, we consider the binary action space for individuals' decision choices, i.e., $A := (\{-1, +1\}^k)$ with k representing the number of tasks an individual is contemplating her action on. We denote the set of k tasks i is pondering actions on by K_i such that $\#K_i = k$. We will consider a richer action space to allow for a continuum $[-1, 1]$ later using the induced norm to define the payoff. These actions are *visible* to other individuals, allowing them to form beliefs about the i's latent characteristics and ideologies. Hence, i's action space is

$$a_{i,l}^t \in \begin{cases} +1 & \textit{if} \text{ individual } i \text{ decides ``yes'' on the item} \\ -1 & \textit{if} \text{ individual } i \text{ decides ``no'' on the item} \end{cases} \tag{3.1}$$

To account for the empirical observation that polarization among the masses takes place at a slower pace and the notion of momentum in individuals' behaviors (Mace et al., 1990), we define the set of individual i's "allowable actions by placing a constraint on the action:

$$A_i^t := \{a_i^t : \left\| a_i^t - a_i^{t-1} \right\| \leqslant \epsilon\} \tag{3.2}$$

Grouping the individual actions for all K tasks, we denote individual i's action profile by $a_i^t := (a_{i,l}^t)_{l \in K_i}$. Similarly, actions for the rest of the society on item K are denoted by $A_{-i,l}^t := (a_{j,l}^t)_{j \neq i, l \in K_i}$ of dimension $(n-1) \times k$, the societal actions by $A_l^t := (a_{j,l}^t)_{j \in [n], l \in K}$, and the actions for those in individual i's neighborhood are expressed $A_{N_i}^t := (a_{j,l}^t)_{j \in [N_i], l \in K_j}$.

3.3 Linear Best Response

We adopt a behavioral model that incorporates three components of players' strategies, and focus on linear best response based on a payoff function, as done in related studies (Bramoullé et al. (2014); Bramoullé and Kranton (2007)). Let $\tilde{a}_i^t$ denote the action that is most *internally* consistent, i.e., with i's ideologies (Y_i^t) and characteristics (X_i^t), which would be the action that i would take in absence of any peer network effects,

$$\tilde{a}_i^t := \max_{a \in \{-1,+1\}^K} \left[\beta_i \langle \psi_i^t \odot Y_i^t, a \rangle + (1-\beta_i) \langle \phi_i^t \odot X_i^t, a \rangle \right] \tag{3.3}$$

with $\beta_i \in [0,1]$ denoting the weight i places on her characteristics versus her ideologies when determining the action. The network at time t is represented by the following adjacency matrix.

$$\mathbf{G}^t := (g_{ij}^t)_{i,j} = \begin{cases} 0 & \textit{if } i = j \text{ or there is no link between individuals } i \text{ and } j \\ 1 & \textit{if } \text{ there is a link between individuals } i \text{ and } j \end{cases}$$

and $\mathbf{G}^0 = (0)_{i,j}$, indicating that there are no social inks at time 0. Together, they form individual i's strategy, which is a mapping of the form

$$s_i : (\{0,1\}) \times [-1,1]^k \times [-1,1]^k \to (g_{ij}^t)_{j \neq i} \times (a_{i,l}^t)_{l \in K_i} \tag{3.4}$$

Note that since s_i is a bounded linear operator, it is continuous[4]. Denote $s := s_i \cup s_{-1}$ and i's strategy space by $\mathcal{S}_i$ and for the whole society by $\mathcal{S} := \times_{i\in[n]}\mathcal{S}_i$.

For any given a^t, each non-elite agent has payoff of the following form:

$$u_i^t(s_i^t, s_{-1}^t) = \alpha_i\big(\langle {A_{N_i}^t}^T \rho_i^t \mathbf{1}, a_i^t\rangle - c_f(deg(i)) - c_a(f_i^t)\big) + (1-\alpha_i)\langle \tilde{a}_i^t, a_i^t\rangle \tag{3.5}$$

where $\alpha \in [0,1]$ is the weight i places on her neighbors' actions ${A_{N_i}^t}^T \rho_i^t \mathbf{1}$, i.e., the peer network effects, and its complement her own ideologies and characteristics. The first term in RHS in (3.5) refers to i's payoff from taking the action similar to those of neighbors', and can thus be called the "peer effect". The term ${A_{N_i}^t}^T \rho_i^t \mathbf{1}$ is commonly referred to as the "local aggregate" in the literature, represeting the weighted average of neighbor's strategies (Parise and Ozdaglar, 2023). The case $\alpha = 0$ corresponds to an agent with no social networks in which case her actions would be determined only by her characteristics and ideologies, while the case of $\alpha = 1$ would correspond to the case of *naive learning agent* whose action is simply a weighted average of her neighbors (Golub and Sadler, 2017). α_i can be interpreted as the extent to which i's action is influenced by affective polarization in that she simply adopts the action most popular in her network.

$(1-\alpha)\langle \tilde{a}_i^t, a_i^t\rangle$ represents i's payoff by acting consistently as dictated by i's characteristics and ideology profiles. $\tilde{a}_i^t := \max_{a\in\{-1,1\}} \big[\beta\langle \phi_i^t \odot Y_i^t, a\rangle + (1-\beta)\langle \psi_i^t \odot X_i^t, a\rangle\big]$ denotes the action that is internally consistent. Following the behavioral economics literature (Akerlof and Dickens, 1982; Tyson, 2008), we say that *cognitive dissonance* stems from an agent choosing an internally inconsistent action, which the agent wishes to minimize, and express it as $\langle \tilde{a}_i^t, a_i^t\rangle$. $\phi_i^t, \psi_i^t, \rho_i^t$ are convex weights. c_a is the peer pressure cost of some action a, increasing in $f_i^t := \left(\frac{\sum_{j\neq i}\langle a_{-jl}^t, a_{il}^t \cdot \mathbf{1}\rangle}{n-1}\right)$, where f_i^t is the "board imbalance" and denotes the fraction of board members that adopts the action opposite of the one taken by agent i.

Let $P_i^t(a_{-i}|X_i^t, Y_i^t, a_{-i}^{t-1})$ be agent i's beliefs about her neighbors' actions at time t and $\Theta(a^t, a^{t-1}, X_i^t, Y_i^t)$ be the probability distribution over a^t induced by P^t. At $t = 0$, we assume

[4] A well-known result in Banach spaces states that for any linear map from a normed linear space to another normed linear space, the following are equivalent about the linear map: (1) it is bounded; (2) it is continuous; and (3) it is continuous at 0. For a detailed discussion, see MacCluer (2009)

$P(1|X_i^0, Y_i^0, \emptyset) = P(-1|X_i^0, Y_i^0, \emptyset) = 1/2$ for $i \in [n]$. Then, agent i's expected utility is

$$\mathbb{E}u_i^t(s_i, s_{-i}) = \sum_{a^t \in N_a^t} \left[\alpha\big(\langle \rho_i^t \odot a_{-i}^t, a_i^t\rangle - c_f(deg(i)) - c_a(f_i^t)\big) + (1-\alpha)\langle \tilde{a}_i^t, a_i^t\rangle\right]\Theta(a^t, a^{t-1}, X_i^t, Y_i^t) \tag{3.6}$$

where the summation is taken over realizations of the other N_a agents' actions in agent i's social network conditional on a_i. This setup induces the game $\Gamma := (N, \{S_i\}_{i\in[N]}, \{u_i\}_{i\in[N]})$.

Solving for the maximum payoff yields the following best response for Stage 1 of the game:

$$f_i(s_{-i}) = \begin{cases} \alpha_i {A_{N_i}}^T \rho_i \mathbf{1} + (1-\alpha_i)\tilde{a}_i & \text{if } \langle {A_{N_i}}^T \rho_i \mathbf{1}, a_i\rangle \geqslant c_f'(deg(i)) + c_a'(f_i) \\ \tilde{a}_i & \text{if } otherwise \end{cases} \tag{3.7}$$

where a_i is the action for i that equates marginal benefits and costs. For Stage 2, i takes a weighted average of the implied action profiles based on the outcome in Equation (3.7).

Due to the Unitary Isomorphism Theorem[5], $\mathcal{H}$ is unitarily equivalent to one based on the space of square-integrable functions on compact Euclidean domains, for instance, $L^2([0,1])$. Hence, we can define our parameters in terms of operators as $\mathbf{A} := [diag((\alpha_i)_i)], \mathbf{P} := [(\rho_{ij})_i], \boldsymbol{a} := (a_i)_i$, and $\tilde{\boldsymbol{a}} := (\tilde{a}_i)_i$, we have

$$(\mathbf{I} - \mathbf{PA})\boldsymbol{a} = (\mathbf{I} - \mathbf{A})\tilde{\boldsymbol{a}} \tag{3.8}$$

Lemma 3.1. There exists an admissible unique equilibrium solution to Stage 1 of the game 3.1.

Proofs are in Appendix. The sketch of the key components of the proof uses the existence of the equilibrium for Stage 2 of the game 3.1, which follows from the convexity of the utility function 3.6 and entails every individual choosing to play an action that is most consistent internally and socially.

[5] For instance, see Corollary 2.5 in Stein and Shakarchi (2003).

3.4 Polarization Measures

Political science literature supports the view that in the first stage of polarization, select few members of the society known as "political elites" become polarized first relatively quickly in the ideologies of the elites. Once ideologies are homogenized, they dictate actions of the elites (Hetherington, 2001; Druckman et al., 2013). In the second stage, the "mass public" comprising the rest of the society becomes polarized, but much more gradually and is characterized as "affective polarization" (Baldassarri and Page, 2021). In our model, the extent of an individual's ideological polarization is parameterized by β and their affective polarization by α. To operationalize α, we define a measure for homogeneity of a neighborhood, which then is used to define a measure of polarization.

Definition 3.2. *The homogeneity h for some neighborhood $\mathcal{N}$ is the map $\mathbb{R} \mapsto [0,1]$ defined $h(\mathcal{N}) := \sum_{i,j\in\mathcal{N}} \big(2 - \|s-t\|\big)/(2\#\mathcal{N})$ where $s,t \in \mathcal{N}$ and $\|\cdot\|$ is the induced L_2 norm in $\mathcal{H}$.*

Homogeneity is defined in terms of the pairwise difference between actions of individuals. We next consider polarization for the smallest group of individuals, a community.

Definition 3.3. *The measure of polarization $\mathcal{Z}$ for any subset $S \in \mathbf{G}$ of the society is the map $\mathbf{G} \to [0,1]$ and defined $\mathcal{Z}(S) := \frac{\sum_{s\in S} h(s)}{\#S}$ where each s is a community in S.*

Then, our proposed measure of polarization for a given subset of the society is the mean of the measures of homogeneity for all communities.

Definition 3.4. *The polarization measure $\mathcal{Z}$ for $\mathcal{S}$ comprising $\#C_S$ communities is the map $\mathcal{S} \mapsto [0,1]$ defined $\mathcal{Z}(S) := \sum_{\forall c,c'\in S} d(c,c')h(s)h(s')/(C_2^S)$ where $c \neq c'$ are communities in $S, d(c,c')$ is the set distance $d(c,c') := \inf(\|x-y\|)$ for any $x \in c, y \in c'$, and $C_2^{\mathcal{S}}$ is the number of combinations of selecting 2 from $\mathcal{S}$.*

The intuition for this definition is that the more polarized a society is, the more similar individuals in each community and more dissimilar individuals in different communities would be.

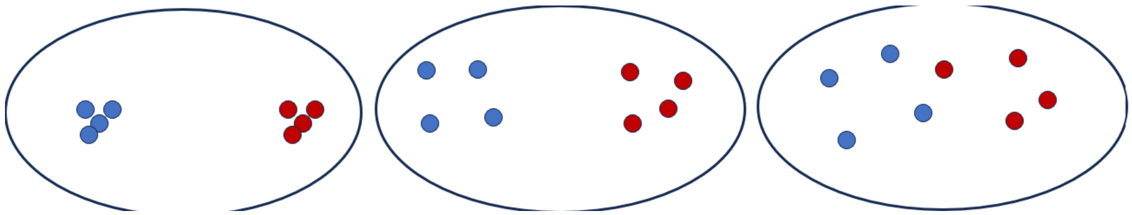

Figure 3.1: The figures above show societies with varying degrees of polarization, starting with most polarized (left) and least polarized (right).

Armed with these two definitions, we operationalize α_i.

Definition 3.5. *The measure of an individual's affective polarization* $\alpha : [0,1] \to [0,1]$ *is a concave function increasing in* $\mathcal{Z}$.

Our interpretation of α as a measure for an individuals' affective polarization finds support in the literature (Baldassarri and Page, 2021). As the communities become more homogeneous, more agents become emotionally attached to neighbors, which influences their ideologies and actions.

4 Equilibrium and Stability in Polarization

We now derive some preliminary results for the society. The first result emphasizes the crucial role that peer effects play in our model in bringing about the "affective polarization" of the masses, and characterizes our setting in terms of complementarity and externality. The former implies that the more of i's neighbors adopt some action, the more i is incentivized to take the same action. The latter implies that the more of an individual's neighbors adopt the same action, the higher the individual's payoff but not necessarily relative to the payoff of not taking the action.

Lemma 4.1. The payoff function in our setting exhibits strategic complementarity and positive externality.

Proofs are in Appendix.

As we will see below, strategic complementarity and positive externality will be crucial in working with a useful tool known as ordinal potential functions, which we use for proofs of some of our key results below. An *ordinal potential function* $f : G(N) \to \mathbb{R}$ for a society with utility function $u := (u_i)_{i\in[n]}$ is one such that g' defeats g if and only if $f(g') > f(g)$ for adjacent graphs

g' and g. A network g' *defeats* an adjacent network g if either (1) $g' = g \setminus ij$ and $u_i(g') > u_i(g)$, or (2) $g' = g + ij$ and $u_i(g') \geqslant u_i(g)$ and $u_j(g') \geqslant u_j(g)$, with at least one being strict inequality and g defeats g' if and only if g is pairwise stable. A sequence of graphs $\{g_1, ..., g_r\}$ with g_j defeating g_i for all $j > i$ is called an *improving path*. Two networks are said to be *adjacent* if they differ by one edge [6]. Our next result relates the notions of stability that we use in the paper.

Lemma 4.2. Let G^{NS} denote the set of Nash stable networks and G^{PS} and G^{PNS} denote the sets of pairwise stable and pairwise Nash stable networks, respectively, in the symmetric network-only setting. Then, $G^{PNS} \subseteq G^{PS} \subsetneq G^{NS}$.

Proofs are in Appendix.

Note that PS networks are not Pareto efficient in the sense that there exists $g \in G, g \neq g_{par}$ such that $u_i(g) \geqslant u_i(g_{par})$ with strict inequality for at least one $i \in [N]$. One can consider a scenario where a single link may not improve an agent's payoff, simultaneous links with many others could. This is due to the positive externality of networks in our setting.

From Equation 3.6, agent i's unique best response is

$$a_i^*(\tilde{a}_i, a_{-i}) = \alpha \rho_i \odot a_{-i} + (1 - \alpha)\tilde{a}_i \tag{4.1}$$

where $\tilde{a}_i$ represents the internally consistent action for agent i. A general expression for all agents can be expressed based on the system of best responses.

4.1 Polarization of the Society

We now study the mechanism by which the larger society becomes polarized, followed by the business community. In rest of this paper, we study pure-strategy equilibria since i's payoff (3.5) is strictly concave in s_i and so there are no mixed-strategy equilibria.

[6] For a detailed discussion of these concepts, see Jackson (2010)

4.1.1 Fixed Ideologies

We consider symmetric games with ideologies fixed at Y_i^0 for all agents. The symmetry implies that network is undirected and induces a network formation game that has been extensively studied in the literature. One variation of the game germane to Stage 1 of our game is known as the link-announcement game (Jackson, 2010), where a link is formed with the voluntary consent of both individuals. In this setting, an individual's decision of whether to add another individual to their network depends solely on the marginal benefit-cost analysis. At time $t = 0$, since the random variables X_{-i}, Y_{-i}^0 are not observable by individual $i, \forall i$, but the actions of other individuals have not been committed yet (i.e., $a_{-i}^0 = \emptyset$), each i bases their action solely on X_i and Y_i^0. Thereafter, others' actions from previous periods are visible, so individuals maintain and initiate links with the others whose expected action matches their own, given the following condition is satisfied.

Lemma 4.3. For $t > 0$, each individual i forms a link with individual j if and only if

$$\langle a_i^t, \rho_j^t p_j^t a_j^t \rangle - c_a(deg(i)) - c_f(f_i^t) > \frac{1-\alpha}{\alpha} \langle \tilde{a}_i^t, a_i^t \rangle \quad \text{subject to } a_i^t \in A_i^t \tag{4.2}$$

where $p_j^t = P(a_j^t | X_i^t, Y_i^t, a_j^{t-1})$.

Proofs are in Appendix.

Remark: The intuition of Lemma 4.3 is that i will form a link with j if (1) j is "sufficiently similar" to i, or (2) she is sufficiently affectively polarized (i.e., large α). Note that for a sufficiently small α, no one would form links with anyone, leading to discrete $\mathbf{G}$ with 0-degree nodes. However, if j is very important to i's decision (i.e., $\rho_j \approx 1$), j's past action predicts his future action perfectly well ($a_j^t = a_j^{t-1}$), and i places equal weight to her neighbors' actions and her own internally consistent action, then the condition simplifies to $\langle a_i^t, a_j^t \rangle - c_a(deg(i)) - c_f(f_i^t) > \langle \tilde{a}_i^t, a_i^t \rangle$, which can never be satisfied as long as either of $c_a(deg(i))$ or $c_f(f_i^t)$ is positive and the starting point is $\tilde{a}_i^t$. But at $t = 0$, the starting point is $\tilde{a}_i^t$ and the two costs take on positive values. This implies no polarization in the case of fixed ideologies, given that the initial distribution of individual ideologies are not multipolar. Hence, there are no disjoint communities and for any pair of agents, there exist lines of communication with the consensus ideology converging to the median of the distribution, i.e., no group polarization in this setup. We formally state this intuition below.

Proposition 4.1. *When individuals' ideologies are fixed according to some initial distribution Y^0, the action profiles and the neighborhoods that form in period $t = 1$ are the equilibrium network and actions, respectively. In particular, if the initial distribution is uniform, then there is exactly 1 connected component in a Nash network in equilibrium, which is also pairwise Nash stable.*

Proof is in the Appendix.

We note that the PNS equilibrium under the assumption of fixed ideologies is highly stable. Regardless of others' actions, an individual's best response is always to act in accordance with her internally consistent $\tilde{a}_i$. In this sense, to describe the equilibrium stability using the notions in Kalai (2020), the index of formation difficulty is 1 and the defection-deterrence index is $|\mathcal{H}|$.

4.2 Variable Ideologies

The main result in this section is that if we allow individuals' ideologies Y to change over time due to interaction with neighbors, then this in turn leads to the polarization and consensus ideologies taking on more extreme values. Recall that individuals' objective is to maximize their expected utility in (3.6), where Y_i^0 is initially assigned by nature according to some symmetric distribution. Consider some set $\mathcal{E}$ of select few "political elites" whose ideologies are polarized[7]. We consider this an exogenous shock to their ideologies at time $t = 1$ such that the implied actions on the action space are lopsided to either unequivocal 1 or unequivocal -1, i.e., $a_i, \forall i \in \mathcal{E} \in \{1^K, -1^K\}$. Members of $\mathcal{E}$ could include, for example, incumbent politicians and political parties.

Definition 4.2. *Political elites are individuals, collectively denoted $\mathcal{E} \subset \mathcal{W}$ and $\#\mathcal{E} \ll \#\Omega$ and $\#\mathcal{E} < \infty$, with the following properties:*

(1) *Their actions are solely driven by their ideologies, and not by their neighbors.*

[7] Such ideologically polarized political elites have been well-documented. See Callander and Carbajal (2022); Krugman (2020)

(2) They are "farsighted" in that their overarching goal is to maximize the size of their neighborhood characterized by uniform actions and ideologies, rather than myopically maximizing 3.6. As a result, they reciprocate all link requests from all $j \in \mathcal{W} \setminus \mathcal{E}$.

(3) They have many more neighbors than non-elites, and much more likely to have neighbors who are also elites.

Property (1) states that elites behave consistently with their ideologies, i.e., their chosen action is the internally consistent $\tilde{a}_i^t$ for all t and $i \in \mathcal{E}$. Property (2) states that in contrast to other individuals who myopically maximize their utility by seeking out like-minded agents, elites have as their ultimate goal convincing other agents to adopt their ideologies through forming extensive relationships. This property implies that c_a is much lower for any $i \in \mathcal{E}$ than those not in $\mathcal{E}$. WLOG, let $c_a = 0$ for any $i \in \mathcal{E}$. Consider the following utility function for political elites:

$$\mathbb{E}u_i^t(s_i, s_{-1}) = \sum_{a_j, j \in N_i} \Big[\alpha(k + \langle a_j^t, a_i^t \rangle) + (1-\alpha)\langle \tilde{a}^t, a_i^t \rangle \Big] \Theta(a^t, a^{t-1}, X_i^t, Y_i^t) \tag{4.3}$$

where $i \in \mathcal{E}$ and the summation is over realizations of the other agents' actions in the network community conditional on a_i. The second term in the square bracket illustrates Property (1) the contribution to utility from taking actions that are internally consistent while the first term represents Properties (2) and (3), and implies that elites prefer to add to their network those who act in similar ways. Property (3) implies that for all $i \in \mathcal{E}, j \notin \mathcal{E}$, we have $deg(i) \gg deg(j)$ and

$$\frac{\#\{j : j \in N_i, j \in \mathcal{E}\}}{\# N_i} \gg \frac{1}{\# N_i} \tag{4.4}$$

As a result, elites form a much larger and more diverse network, i.e., $\alpha_i \ll \alpha_j, c_{i,a} \ll c_{j,a}, \forall i \in \mathcal{E}, j \notin \mathcal{E}$, and their degrees are much larger than the mean of their neighbors $deg_i \gg \mathbb{E}_{j \in N_i}(deg_j)$. For concrete examples, one could think of Congresswoman Alexandra Ocasio-Cortez of New York, Governor Ron DeSantis of Florida, and former President Donald Trump. Political elites are characterized by several distinguishing features.

4.2.1 Elite Polarization

To make this example more concrete, suppose that some $2m \ll n$ elites are exogenously polarized, resulting in m of them adopting the action 1_K and the other m adopting -1_K. In other words, we have $a_i^0 = a_i^t = 1_k$ for $i \in [m]$ and $a_i^0 = a_j^t = -1_k$ for $j \in [m], t \leqslant T$. Consider agent i who is not an elite but has an elite as their neighbor, and we would like to investigate the influence that this elite has on i's behavior. Let $a_{e,k}$ denote the elite's action on some issue $k \in [K]$. Clearly, if $a_{e,l} = \tilde{a}_{il}$, then acting in unison with the elite would yield the payout $\beta\phi_{ik}\psi_{il} + (1-\beta)\rho_{ie}$. In general, non-elites' ideology evolves due to interactions with their neighbors as follows:

$$Y_i^t = Y_i^{t-1} + \delta\Big((a_{N_i}^t)^T \rho_i^t - Y_i^{t-1}\Big), \delta \in (0,1) \tag{4.5}$$

We now look at under what conditions would i change their decision to match that of the elite.

Lemma 4.4. Let $a_i^* \in A_i^t$ denote the action that maximizes the utility in (3.5) in equilibrium. A non-elite individual i forms a link with an elite $e \in \mathcal{E}$ if and only if the following condition holds:

$$\langle a_e^t, a_i^* - a_i^t \rangle > \frac{c_f'(deg_i^t) + c_a'(f_i^t) + (1-\alpha_i)\langle \tilde{a}_i^t, a_i^* - a_i^t \rangle}{\alpha_i \rho_{i,e}^t} \tag{4.6}$$

That is, we must have (a) i's action is sufficiently homogeneous, (b) society is not too unbalanced ideologically, (c) c_f must be sufficiently concave, (d) they place sufficient weight on elites in their network, and (f) they must be sufficiently affectively polarized.

Proof is in the Appendix. Result below states the condition for mass polarization.

Lemma 4.5. Individual i's internally consistent action becomes more polarized if their ideology changes enough in a time increment. That is, let J be the set of indices for the elements in Y_i^{t-1} such that for any $j \in J$, $sgn(Y_{ij}^{t-1}) \neq sgn\big(\langle Y_i^{t-1}, 1\rangle\big)$. Then,

$$\delta > \min_j \frac{Y_{ij}^{t-1}}{\big[(a_{N_i}^t)^T \rho_i^t\big]_j - Y_{ij}^{t-1}} \tag{4.7}$$

Proof is in Appendix.

Our main result for this section is given below as Proposition 4.3, which states that assuming evolving ideologies as in Lemma 4.3 and equal-distributed elite executives as hypothesized above, then a unique PNS network results with two disparate components with polarized action profiles. A sketch of the proof for this statement goes as follows. The previous two lemmas give us that initially, individuals with more homogeneous action profiles form links with the elites and consequently become polarized in each time step. Then, action profiles of the individuals' neighbors in their networks become more homogeneous, which in turn make their ideologies more homogeneous. Since in each time step, at least one individual's action and ideology profiles become more homogeneous, and N_e is finite, every individual's action profile will be either $+1_K$ or -1_K, which rules out links between agents with different action profiles. Once everyone is polarized, it become apparent that the network is both NS and PS, and thus PNS by Lemma 4.2.

Proposition 4.3. *Suppose some fixed number m of elite executives experience exogenous bipolar shocks to their ideologies akin to the "Democrat vs. Republican" affective political polarization, and are characterized by the utility function 4.3. Then, within a finite number of time iterations, there will be exactly 2 communities that are disconnected from each other in a Nash network in a sub-game perfect equilibrium.*

In contrast to the case of fixed ideologies in §4.1.1, the PNS equilibrium under variable ideologies is less stable. To see why, take the bipolar community result in Proposition 4.3 and notice that the size of either community is $|\mathcal{H}|/2$. Since the game exhibits strategic complementarity, in the case of finite population, a defection of any neighbor would reduce the payoff of staying in the community. This implies that the defection-deterrence index is 1. At the same time, this equilibrium is relatively more difficult to form compared to the fixed ideology case, since any individual would need half the population to join to make it worthwhile to stay in that community. This implies the index of formation difficulty is $|\mathcal{H}|/2$.

5 Auditor Screening

We now investigate the problem of screening auditors from the perspective of audit committee of a corporate board of directors at a given time t and its state of the polarization $\mathcal{Z}(\mathcal{S}_t)$. A pool of auditors comprises a finite sample from the society $\mathcal{S}_t$ and are referred to as "agents" and the impartial audit committee is the principal. As a sample from the population, agents inherit different values of polarization parameters (e.g., the social polarization parameter α_i^t) that have been determined in the broader society at a given phase of polarization. Let $\mathcal{A} \subset \mathbf{V}, \#\mathcal{A} = m$, denote the finite set of indices for the individuals in the society each of whom is a Certified Public Accountant (CPA) auditor. If there was a single auditor agent under consideration by the committee, then one would proceed with the standard theory of contracts to manage adverse selection inherent in such selection processes. However, in our setting, there are multiple $n > 1$ auditors vying for the audit engagement for the committee, giving rise to each agent forming beliefs about other agents and taking strategic actions in response. To model this situation, we use analytical tools from the mechanism design literature.

At a given time period t, an audit committee of a corporate board of directors is tasked with selecting an auditor from $\mathcal{A}$ for the upcoming fiscal year. The committee's objective is to maximize the veracity of the audited financial report which we take to be increasing in the number and quality of bids received, and is hence risk-neutral. Auditors have additively separable von Neumann-Morgenstern utility that takes the value zero if they are not chosen by the committee and the value $\theta_i - c_i$ if they are chosen, where $\theta_i \in [0, 1]$ is auditor i's type and c is the cost of participating in the "pitch" during the selection process. Using this notation, the audit committee's objective is to maximize its total "payments" received of $\sum_i c_i$. As we will see, auditor's type θ_i is a random variable and can be thought of as i's measure of social polarization relative to the CFO that determines the probability of their selection in the screening process. It is also their valuation of the audit engagement since they are indifferent between paying θ and getting the engagement awarded and not getting selected. On the other hand, c could incorporate costs such as the junior staff's time in preparing pitching materials and the relationship partners' time to pitch the audit firm's qualifications.

Specific value of θ_i (i.e., the auditor's polarization measure) is private to the auditor and is independent of any other θ_j for all $j \neq i$. We collectively refer to the types of all agents other than i by θ_{-i}. The audit committee forms a belief about θ_i's in the form of $\Delta(\theta_i)$, which denotes the set of all probability distributions over θ_i, which in turn is dictated by cumulative probability distribution F_i conditional on the polarization measure $\mathcal{Z}(\mathcal{S}_t)$ and its density f_i with support $[\underline{\theta}, \overline{\theta}]$, where i and j may have different distributions. The cdf and pdf of θ_{-i} are denoted by F_{-i} and f_{-i} with the support $\Theta_{-i} := [\underline{\theta}, \overline{\theta}]^{m-1}$. The cdf for all auditors jointly is denoted $F := F_1 \times F_2 \times \cdots \times F_m$ and similarly for the density f. The support for the cdf and the pdf is $[\underline{\theta}, \overline{\theta}]^m$. Although specific realizations of θ_is are independent, they are generated from a common prior F and this is common knowledge among all participants in the screening problem. Note that all probability distributions are conditional on a given polarization measure $\mathcal{Z}(\mathcal{S}_t)$ defined in (3.4). Rather than denoting by $F_i(\cdot : \mathcal{Z}(\mathcal{S}_t))$, for notational simplicity, we use the shorthand F_i^t or when t is understood as the conditional variable, we drop it altogether and just say F_i and similarly for f_i and F.

We provide some remarks on our setup above. The assumption that each agent's type realizations are independent is based on the observation that auditors in the pool can come from any geographical area, and the fact that one auditor is based on in area does not affect the probability of another agent's coming from any specific area. The assumption that auditors' private types are generated from the common distribution F and that it is a common knowledge is reasonable given that the population-level polarization measures are common to all auditors based in close geographical proximity. For instance, if all auditors in the pool are from a same geographical area, then the underlying cdfs would be the same. The virtual valuation is defined:

$$\psi_i(\theta_i) = \theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)}$$

We want to characterize mechanisms for selecting an auditor that maximizes the audit committee's utility. The audit committee's choice variables are extensive games and strategies in those games subject to the constraints of individual rationality (IR) and incentive-compatibility (IC) on the part of auditors. To each leaf in the extensive game tree, the audit committee assigns a probability using an “allocation rule" $q(\theta) := (q_1(\theta), \cdots, q_m(\theta))$ to each auditor of getting the engagement awarded, where $q_i(\theta)$ is the probability that auditor i is awarded the engagement conditional on the type vector $\theta = (\theta_1, \cdots, \theta_m)$. This defines a game of incomplete information and we adopt the standard solution concept that is used in such games in the literature, Bayesian Nash Equilibrium (BNE).

Since the audit committee can completely commit to the selection procedure, we can appeal to the Revelation Principle and without loss of generality limit our investigation of screening procedures to direct mechanisms to characterize optimal auditor selection procedure in terms of allocation rule and payment amount. A direct mechanism is chosen by the audit committee and auditors truthfully report their types in the extensive BNE game chosen by the committee subject to IR and IC. For any given direct mechanism, define the conditional expectation of auditor i winning the bid:

$$Q_i(\theta_i) = \int_{\Theta_{-i}} q_i(\theta_i, \theta_{-i}) f_{-i}(\theta_{-i}) d\theta_{-i}$$

and the conditional expectation of auditor i's transfer payment to the audit committee

$$T_i(\theta_i) = \int_{\Theta_{-i}} t_i(\theta_i, \theta_{-i}) f_{-i}(\theta_{-i}) d\theta_{-i}$$

Then, the necessary and sufficient conditions for a direct mechanism to be IC are that Q_i is increasing and $T_i(\underline{\theta}_i) \leqslant \underline{\theta}_i Q_i(\underline{\theta}_i)$. Due to Myerson (1981), we show below that the audit committee can simply pick the following allocation rule to ensure that its mechanism meets the IC and IR constraints while optimizing its payoff:

Theorem 5.1. *Suppose F_i for all i are regular. Define the agent type θ_i as follows:*

$$\theta_i = \alpha_i \cdot h(N_i) \cdot \mathbf{1}(C_i \neq C_{CFO})$$

where α_i is i's social polarization parameter, $h(N_i)$ is the homogeneity measure of the auditor's social network, and the last term is the Kronecker's delta that evaluates to 1 if the community of the auditor is different from that of the CFO. Then the mechanism that maximizes the audit committee's expected payoff has

$$q_i(\theta) = \begin{cases} 1 & \text{if } \psi_i(\theta_i) > 0, \psi_i(\theta) > \psi_j(\theta_j) \forall j \neq i \\ 0 & \text{if otherwise} \end{cases}$$

and

$$T_i(\theta_i) = \theta_i Q_i(\theta_i) - \int_{\underline{\theta}}^{\theta_i} Q_i(x)dx$$

6 Discussion and Conclusion

In the current politically polarized environment, auditor's decision-making on audit engagements is increasingly influenced by emotional attachments to their polarizing social networks vis-a-vis that of the CFO. Selecting auditors whose social polarization is as far removed as possible from the CFO would optimize the audit committee's payoff in terms of veracity and quality of the audited financial statements. While audit committee was assumed to be impartial, investigating how polarization affects audit committee's ability to meet different types of obligations would be promising future directions for research.

7 Appendix

7.1 Proofs

Proof. of Lemma 3.1

The solution to Stage 1 of the game 3.1 is obtained by solving (3.8) for $\boldsymbol{a}$, which presents a unique solution since the operator $(\mathbf{I} - \mathbf{PA})$ is invertible. The admissible part is shown by noting that the ratio of the minimum eigenvalue of $(\mathbf{I} - \mathbf{PA})$ and the maximum eigenvalue of $(\mathbf{I} - \mathbf{A})$ is less than 1, which implies that the best response operator is a contraction. This, together with the continuously differentiability of the payoff function and the fixed point theorems, yields the uniqueness as desired. ■

Proof. of Lemma 4.1 WLOG suppose $a_i = 1_K$. Examining the expression for the payoff function (1a), it is clear that $u_i(a_i, a_{-i})$ increases as $\langle a_i, a_{-i} \rangle$ increases, proving the claim about positive externality. Further, take WLOG a'_i to be the same as a_i except replace some j-th entry with 0. Then, notice that $u_i(a_i, a_{-i}) - u_i(a'_i, a_{-i}) = K > u_i(a_i, a'_{-i}) - u_i(a'_i, a'_{-i}) = K - 1$, thus proving the claim about strategic complementarity. ■

Proof. of Lemma 4.2 We first show that a pairwise stable network in our setting is Nash stable. Let $g \in G^{NS}$. Then, for all pairs of agents, at least one agent is worse off from adding an edge and neither is better off from removing an edge. Clearly, no agent can unilaterally improve their payoff, and this gives us the Nash equilibrium. For an example of a Nash equilibrium that is not pairwise equilibrium, consider a network with no edges where a pair of agents could benefit with a positive payoff with a link. The first set inclusion follows from definition of pairwise Nash stability. ■

Proof. of Lemma 4.2 The hypothesis is equivalent to requiring that the net payoff from forming the link with j outweighs the cognitive benefit of acting consistently with i's characteristics and ideologies, normalized by the affective polarization measure, and rearranging terms. ■

Proof. of Lemma 4.1 Recall that Y^0 is assumed to be distributed symmetrically. This implies that for every i, $\#N_i := \#\{j | j > \frac{c}{(1-\alpha)(\rho_j p_j)}\}$ is the same for all i. No agent can improve their utility by changing their behavior on any of the K policies since the distribution of a_i is uniform. Let $C \in g$ be the connected component, i.e., a neighborhood, that results and define $f(C) = \sum_{i,j \in C} \langle a_i, a_j \rangle$. We want to show that f is an ordinal potential function for C. Take any network C' that is adjacent to C. By Lemma 4.3, if $C' = C - ij$, then $u_i(C) > u_i(C')$ as well as if $C' = C + ij$, then $u_i(C) \geqslant u_i(C')$ and $u_j(C) \geqslant u_j(C')$. By the properties of ordinal functions, the connected component is pairwise stable, and by Lemma 4.2, it is also pairwise Nash stable. ■

Proof. of Lemma 4.4 Recall that $a_e^t \in \{1_k, -1_k\}$. Hence, maximizing the LHS of (3) necessitates that a_i^* be as homogeneous as possible. Also, recall that c_a and c_f are convave and increasing in f and deg_i, respectively. Individual i placing sufficient weight on the political elites in their network is tantamount to $\rho_{i,e}$ being sufficiently high. Affective polarization is directly measured by α_i. These amount to decreasing the quantity in the RHS of (3). ■

Proof. of Lemma 4.5 Making $\tilde{a}_i$ more homogeneous entails finding the more common element in Y_i^{t-1}, i.e., $sgn(\langle Y_i^{t-1}, 1\rangle)$. WLOG suppose this element is +1. Then, we would like δ to be just big enough to make the negative element with the least hurdle positive. The closed-form expression for this is precisely (4). ■

Proof. of Proposition 4.3 To see how introducing polarized elites and letting ideologies evolve lead to two disparate subnetworks, assume first that society is not too ideologically unbalanced. Initially, some i whose action is sufficiently extreme forms link with $G^{\mathcal{E}}$. Then, i's ideology becomes more extreme by the amount $\delta\bigg(\left(a_{N_i}^t)\right)^T \rho_i^t - Y_i^{t-1}\bigg)$. For a sufficiently large δ, this leads to more extreme $\tilde{a}_i^t$ and hence a_i^1. Then, more elite executives discover agent i, forming links more easily and increasing $\rho_G^{\mathcal{E}}$. This in turn makes $a_j^1, \forall j \in N_i^0$ more extreme. So, in each step, some agents' action becomes more polarized, and this continues for a finite number of iterations since $n, k < \infty$. Now, we present the proof of the statement.

When agent j who is not an elite comes in contact with an elite i, two things happen. First, by being in j's network, the elite's action has an immediate effect on j's action decision. More importantly, through frequent interactions, persuasive rhetoric and correspondence, the weight that j places on i as captured by ρ_{ij} will be high, making elite i's influence on j's action stronger. For those individuals who base their decision more on their neighbors' actions than on their ideologies and circumstances, i.e., those with relatively lower values of α, the change in their behaviors would be immediate. On the other hand, those with a stronger basis on their ideologies, i.e., those with higher α values would experience the cognitive dissonance arising from the mismatch between the action implied by their ideologies and circumstances, and that implied by the actions from the political elites in their networks. Allowing the ideologies of these individuals to slowly change, we can see that they would be able to increase their utility by adjusting their ideologies such that the implied action more closely matches that implied by their neighbors, in particular the elites. ■

Proof. of Proposition 5.1 Note that θ is a random variable taking values in the closed unit interval. The rest follows from the strict application of the Myerson's lemma on revenue-maximizing auction in Myerson (1981). ■

Bibliography

Daron Acemoğlu, Giacomo Como, Fabio Fagnani, and Asuman Ozdaglar. Opinion fluctuations and disagreement in social networks. *Mathematics of Operations Research*, 38(1):1–27, 2013.

George A Akerlof and William T Dickens. The economic consequences of cognitive dissonance. *The American economic review*, 72(3):307–319, 1982.

Gene Ambrocio and Iftekhar Hasan. Belief polarization and covid-19. *Bank of Finland Research Discussion Paper*, 1(10), 2022.

Mazhar Arikan, Mehmet Kara, Adi Masli, and Yaoyi Xi. Political euphoria and corporate disclosures: An investigation of ceo partisan alignment with the president of the united states. *Journal of Accounting and Economics*, page 101552, 2022.

Delia Baldassarri and Scott E Page. The emergence and perils of polarization. *Proceedings of the National Academy of Sciences*, 118(50):e2116863118, 2021.

Theodore Bergstrom, Lawrence Blume, and Hal Varian. On the private provision of public goods. *Journal of public economics*, 29(1):25–49, 1986.

Avishek Bhandari, Joanna Golden, and Maya Thevenot. Ceo political ideologies and auditor-client contracting. *Journal of Accounting and Public Policy*, 39(5):106755, 2020.

Ugo Bolletta and Paolo Pin. Polarization when people choose their peers. *Available at SSRN 3245800*, 2020.

Narjess Boubakri, Omrane Guedhami, Dev Mishra, and Walid Saffar. Political connections and the cost of equity capital. *Journal of corporate finance*, 18(3):541–559, 2012.

Yann Bramoullé and Rachel Kranton. Public goods in networks. *Journal of Economic theory*, 135 (1):478–494, 2007.

Yann Bramoullé, Rachel Kranton, and Martin D'amours. Strategic interaction and networks. *American Economic Review*, 104(3):898–930, 2014.

Steven Callander and Juan Carlos Carbajal. Cause and effect in political polarization: A dynamic analysis. *Journal of Political Economy*, 130(4):825–880, 2022.

Harold Clarke, Marianne C Stewart, and Karl Ho. Did covid-19 kill trump politically? the pandemic and voting in the 2020 presidential election. *Social Science Quarterly*, 102(5):2194–2209, 2021.

Geoffrey L Cohen. Party over policy: The dominating impact of group influence on political beliefs. *Journal of Personality and Social Psychology*, 85(5):808, 2003.

Nelson Cowan, Emily M Elliott, J Scott Saults, Candice C Morey, Sam Mattox, Anna Hismjatullina, and Andrew RA Conway. On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive psychology*, 51(1):42–100, 2005.

Nicholas Dias and Yphtach Lelkes. The nature of affective polarization: Disentangling policy disagreement from partisan identity. *American Journal of Political Science*, 66(3):775–790, 2022.

James N Druckman, Erik Peterson, and Rune Slothuus. How elite partisan polarization affects public opinion formation. *American Political Science Review*, 107(1):57–79, 2013.

Adam M Enders. Issues versus affect: How do elite and mass polarization compare? *The Journal of Politics*, 83(4):1872–1877, 2021.

Robert Felix, Sattar Mansi, and Mikhail Pevzner. Audit committee–cfo political dissimilarity and financial reporting quality. *Journal of Accounting and Public Policy*, 45:107209, 2024.

Vyacheslav Fos, Elisabeth Kempf, and Margarita Tsoutsoura. The political polarization of corporate america. Technical report, National Bureau of Economic Research, 2022.

Andrea Galeotti, Sanjeev Goyal, Matthew O Jackson, Fernando Vega-Redondo, and Leeat Yariv. Network games. *The review of economic studies*, 77(1):218–244, 2010.

Benjamin Golub and Evan Sadler. Learning in social networks. *Available at SSRN 2919146*, 2017.

Mostafa Monzur Hasan and Haiyan Jiang. Political sentiment and corporate social responsibility. *The British Accounting Review*, page 101170, 2022.

Marc J Hetherington. Resurgent mass partisanship: The role of elite polarization. *American political Science Review*, 95(3):619–631, 2001.

Matthew O Jackson. Social and economic networks. In *Social and Economic Networks*. Princeton university press, 2010.

Sham Kakade, Michael Kearns, John Langford, and Luis Ortiz. Correlated equilibria in graphical games. In *Proceedings of the 4th ACM Conference on Electronic Commerce*, pages 42–47, 2003.

Adam Tauman Kalai and Ehud Kalai. Best-response reasoning leads to criticalmass equilibria. *Preprint*, 2021.

Ehud Kalai. Viable nash equilibria: Formation and defection. *Preprint, Northwestern University, February*, 2020.

Michael Kearns, Michael L Littman, and Satinder Singh. Graphical models for game theory. *arXiv preprint arXiv:1301.2281*, 2013.

Duk Gyoo Kim, Daehong Min, and John Wooders. Viable nash equilibria: An experiment. *CESifo Working Paper*, 1(9913), 2022.

Paul Krugman. *Arguing with zombies: Economics, politics, and the fight for a better future*. WW Norton & Company, 2020.

Geoffrey C Layman, Thomas M Carsey, and Juliana Menasce Horowitz. Party polarization in american politics: Characteristics, causes, and consequences. *Annu. Rev. Polit. Sci.*, 9:83–110, 2006.

Barbara D MacCluer. *Elementary functional analysis*. Springer, 2009.

F Charles Mace, Joseph S Lalli, Michael C Shea, Elizabeth Pinter Lalli, Barbara J West, Maura Roberts, and John A Nevin. The momentum of human behavior in a natural setting. *Journal of the Experimental Analysis of Behavior*, 54(3):163–172, 1990.

Lilliana Mason. *Uncivil agreement: How politics became our identity*. University of Chicago Press, 2018.

Jennifer McCoy, Tahmina Rahman, and Murat Somer. Polarization and the global crisis of democracy: Common patterns, dynamics, and pernicious consequences for democratic polities. *American Behavioral Scientist*, 62(1):16–42, 2018.

Roger B Myerson. Optimal auction design. *Mathematics of operations research*, 6(1):58–73, 1981.

Olena Orlova. Idiosyncratic preferences in games on networks. *Games and Economic Behavior*, 131:29–50, 2022.

Francesca Parise and Asuman Ozdaglar. Graphon games: a statistical framework for network games and interventions. *Econometrica*, 91(1):191–225, 2023.

Elias M Stein and Rami Shakarchi. *Princeton lectures in analysis*. Princeton University Press Princeton, 2003.

Alexander G Theodoridis. Me, myself, and (i),(d), or (r)? partisanship and political cognition through the lens of implicit identity. *The Journal of Politics*, 79(4):1253–1267, 2017.

Amos Tversky and Daniel Kahneman. The framing of decisions and the psychology of choice. *science*, 211(4481):453–458, 1981.

Christopher J Tyson. Cognitive constraints, contraction consistency, and the satisficing criterion. *Journal of Economic Theory*, 138(1):51–70, 2008.

Paul Whitney, Christa A Rinehart, and John M Hinson. Framing effects under cognitive load: The role of working memory in risky decisions. *Psychonomic bulletin & review*, 15(6):1179–1184, 2008.